Vol.2009-MUS-80 No.7
2009/5/21

楽曲類似検索のための圧縮オーディオファイル形式からの高効率特徴抽出手法

青木圭子　　神田龍一　　帆足啓一郎　　柳原広昌†

楽曲の類似検索における高速化の検討と実装について報告する．楽曲の特徴量抽出では，PCMのデータから係数抽出を行うMFCC係数が一般的に用いられている．本稿では，楽曲の圧縮ファイルから直接MFCCに近い係数を抽出し，高速化する手法を提案する．本手法をシステムに実装した結果，従来手法の約30倍に高速化することができた．

Efficient Feature Extraction from Compressed Audio Files for Content-based Similar Music Retrieval

Keiko Aoki　Ryuichi Kanda　Keiichiro Hoashi　and Hiromasa Yanagihara†

This paper proposes an efficient feature extraction method from compressed audio files for content-based similar music retrieval. The proposed method improves the extraction speed by extracting audio features directly from the compressed domain from audio files. We have conducted experiments to evaluate the efficiency and accuracy of the proposed method, and the results indicate that our method is capable of extracting features 30 times faster than the conventional method, while preserving accuracy of music retrieval.

1. はじめに

楽曲の類似検索においては，楽曲ファイルの音響的特徴量を量子化し，コサイン距離等でその類似度を測定する手法が用いられている．中でも音声認識のモデル学習で知られるMFCC(Mel-Frequency Cepstrum Coefficient)が特徴量として広く用いられている［1］．

一方，昨今の音響圧縮技術の発達により，楽曲データは，MP3(Mpeg-1 Audio Layer-3)[2]やAAC(Advanced Audio Coding)[3]等の圧縮形式で保存することが多くなった．MFCCを求める場合には，これらの圧縮ファイルを一旦非圧縮のPCM形式にデコードした後，特徴量の計算を行う必要がある．そのため，楽曲の類似検索においては，特徴量抽出過程がシステム構築にかかる時間の大半を占めている．

そこで筆者らは，携帯端末を対象とした音楽配信サービスにおいて主流となりつつある，HE-AAC(High-Efficiency Advanced Audio Coding)の楽曲圧縮ファイルから直接MFCCに相当する特徴量（AACCEP）抽出を行う手法[4]を提案した．本稿では，さらに精度向上させるため，MFCCで導入されている高域強調処理も考慮した改善方式を提案し，他の手法との比較も含め，実装・評価を行った．

2. 従来手法

2.1 MFCC

従来のMFCCの特徴量抽出の手順は図 1の通りである．PCMデータを入力として高域強調を行い，フーリエ変換を行った後，メル周波数での帯域分割を行い，DCT変換を行う．

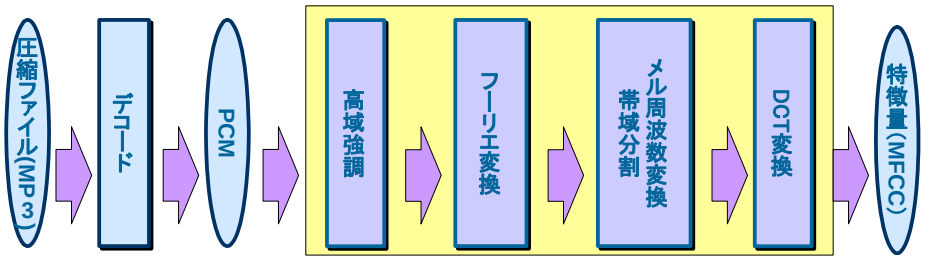


図 1　MFCC の処理手順

† KDDI 研究所
KDDI R&D Laboratories Inc.

MFCC の特徴量抽出処理においては，上記手順のうち，フーリエ変換部分に約 75%の処理時間が費やされている．

2.2 MP3CEP

MP3 データから直接特徴量抽出を行う手法として，MP3CEP[5]が提案されている．MP3CEPはMP3 データをフィルタバンク出力部分までデコードし，その各サブバンドデータの対数にDCT変換を行うことで特徴量算出を行う手法である（図 2）．

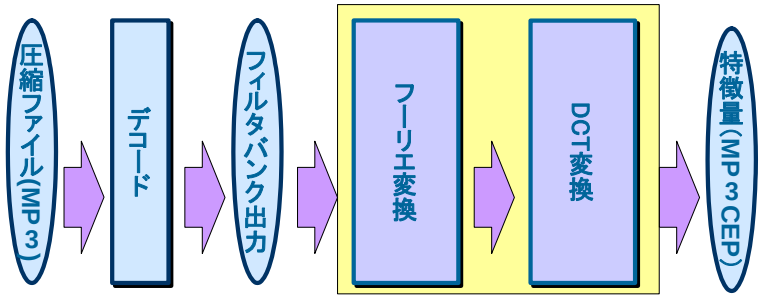


図 2　MP3CEP の処理手順

特徴量抽出手順で使用するデータは，PCMデータではなく，MP3 のデコード中に得られるフィルタバンク出力である．MP3 のデコードでは，最後にフィルタバンクの合成を行っているため，合成前のフィルタバンクのデータを用いる（図 3）．MP3 符号化では一旦，時間領域のフィルタで 32 サブバンドに分割した後にMDCTを行うのに対し，HE-AACでは入力サンプルに直接MDCTが行われるため，本手法は適用できない．

3. 提案手法

3.1 AACCEP（改善提案方式）の概要

[4]において提案したHE-AACデータからの高速特徴量抽出方式を拡張し，MFCC処理においてスペクトラム平坦化によるSNR向上を目的に導入されている高域強調処理を考慮した改善方式を提案する．本提案方式の処理手順を図 4に示す．今回，HE-AACデータ（44.1kHz)のSBR成分については考慮しないため，AACデコード処理の途中で得られる半分の周波数（22.05kHz）に相当するMDCT係数（1024 個/フレーム）を取り出し，周波数ドメイン上で高域強調フィルタを掛ける（詳細は後述）．その後，人間の音の高低に対する聴覚特性を反映したメルフィルタバンクに写像して 26 個の係数（詳細は後述）を得た後，DCT変換を行ったものから低域の 12 係数を取り出したものを本提案手法における特徴量（AACCEP）とする．ここでMFCCにおけるフーリエ変換で得られる値の代りに，AACのMDCT係数を用いた理由としては，共にほぼ同一レンジ（20msec前後/フレーム）での周波数特性を表現するものであることと，AACデコード処理において大半の処理時間が費やされているIMDCT処理の前段階でMDCT係数を抽出することで，大幅な時間短縮が可能であることが挙げられる（図 5）．尚，MSステレオ, インテンシティステレオ, TNSの処理はデータ形式によっては必須ではない．

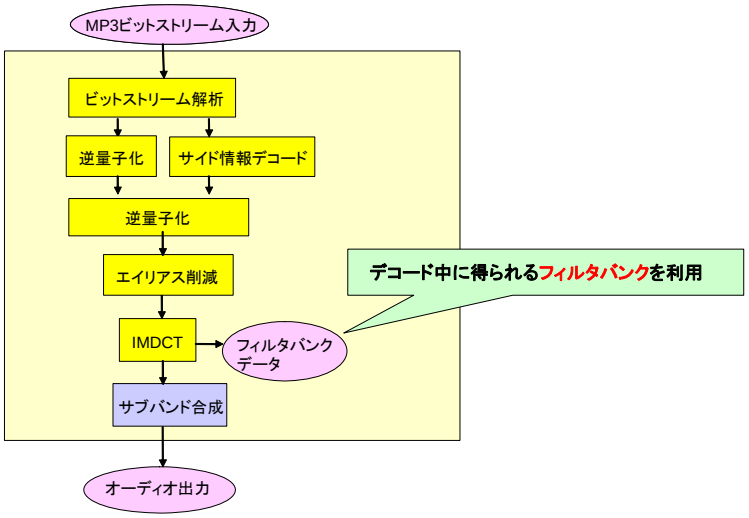


図 3　MP3 のデコード手順

Vol.2009-MUS-80 No.7
2009/5/21

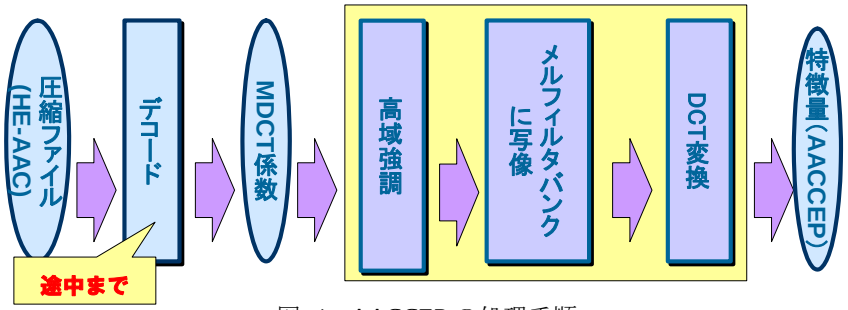


図 4　AACCEP の処理手順

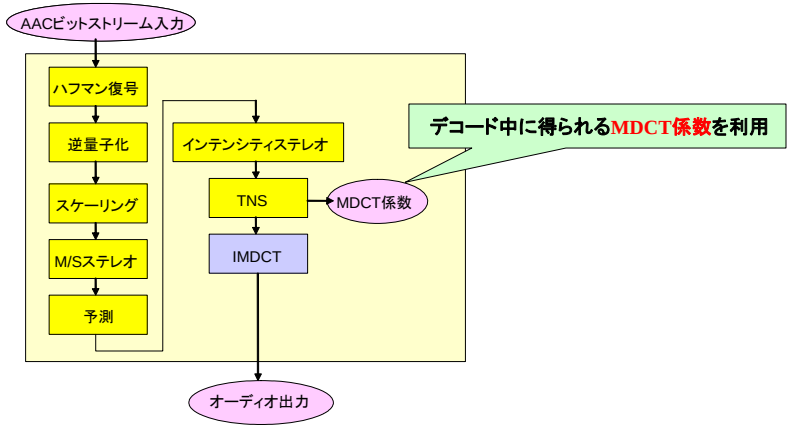


図 5　HE-AAC のデコード手順

3.2　高域強調

MFCC の特徴量抽出処理においては，入力データに対して[-0.97，1.0]の移動平均フィルタを用いた高域強調処理が行われている．MFCC の値に近づけるため，AACCEP においてもフィルタ関数による高域強調処理を取り入れる．MFCC では時間軸上で移動平均フィルタをかけているが，AACCEP では周波数軸上で実行するため，移動平均フィルタ H(z)を

$$H(z) = -0.97z^{-1} + z^{0} \qquad \text{（数式 1）}$$

と定義し，その周波数特性 H(f)を求めた（数式 2）．

$$|H(f)| = \sqrt{1.9409 - 1.94\cos 2\pi f T_s} \qquad \text{（数式 2）}$$

T_s:標本化周期

移動平均フィルタの周波数特性を図 6に示す．メルフィルタバンク処理の前にこの値を乗ずることで高域強調を行った．

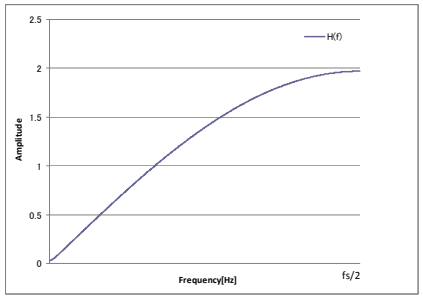
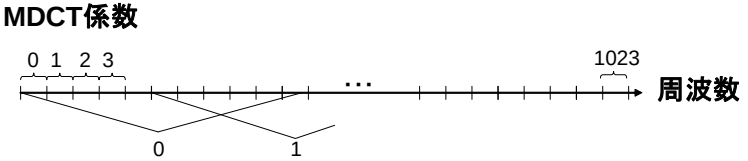


図 6　移動平均フィルタの周波数特性

3.3　メルフィルタバンクへの写像

MDCT係数からメルフィルタバンクへの写像方法を説明する（図 7）．

メル周波数の各フィルタバンクを，MDCTで用いる通常の周波数スケールに変換し，その間に含まれるMDCT係数の絶対値に窓関数（図 8）を掛けて加算することで写像を行っている．その後，生成された 26 個の係数に対してDCT変換を行ったものから取り出した低域の 12 係数を各メルフィルタバンクにおける係数（AACCEP）とする．



メル周波数=2595･log10(1+周波数/700)

メルフィルタバンク	0	1	2	・・・	24	25
MDCT 係数範囲	0～10	4～19	11～29	・・・	765～966	861～1023

図 7　メルフィルタバンクへの写像手順

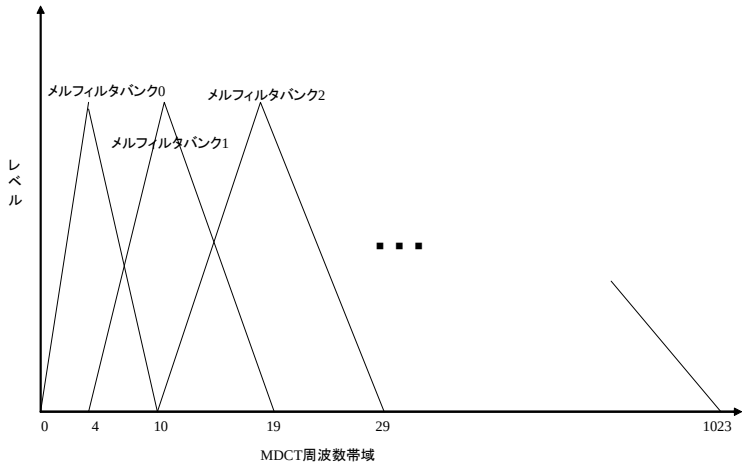


図 8　メルフィルタ処理における窓関数の例

4.　実験

本提案手法の効果を検索速度と検索精度の両面で比較検証を行った．

4.1　システム構成

本実験を行うために楽曲類似検索システムを構築した．その構成を図 9に示す．

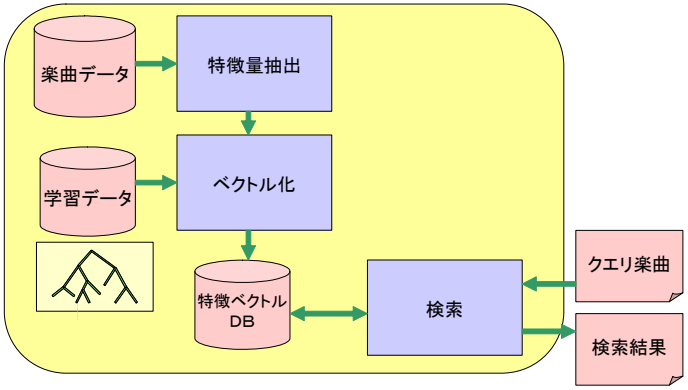


図 9　システム構成

具体的なシステムの構築手順と楽曲の類似検索の実施手順を以下に示す（図 10）．

(1)　特徴空間の作成

１．　学習データ（ジャンル情報）からTreeQ[6]を使ってツリーを作成
２．　全楽曲データについて，フレーム毎の特徴量（MFCC etc.）を抽出
３．　１で作成したツリーを使ってヒストグラムを作成（ベクトル化）

３で作成したベクトルを特徴ベクトルとした特徴空間が作成される．特徴空間作成処理においては２の特徴量抽出処理の負荷が特に高い．

(2)　類似検索

(1)で作成した特徴空間において，クエリ楽曲と検索対象楽曲とのコサイン距離を求め，距離が近い楽曲を類似度の高い楽曲とすることで検索する．さらにクラスタリングを行い，クラスタ単位の学習データを抽出することでツリーの最適化が可能である．

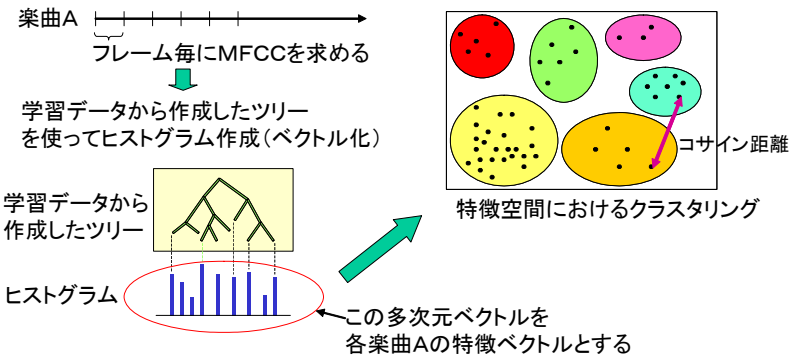


図 10　システム構築手順

4.2　速度検証

(1)　検証条件

本稿では，表 1の実行環境とデータを用いて検索速度の検証を行った．100 曲の楽曲について特徴量抽出を行い，その平均実行時間を求めた．

Vol.2009-MUS-80 No.7
2009/5/21

実行環境		データ	
マシン	Endeavor MT7800	ファイル数	100曲
CPU	Core2 Duo E6700 2.66GHz	形式	HE-AAC、MP3
OS	Vine Linux 4.1	サンプリング周波数	44.1kHz
メモリ	3GB	ビットレート	48kbps

表 1　実行環境とデータ

(2) **検証結果**

検証の結果，提案手法を用いた場合，MFCCに比べ，3.3%の時間（約 30 倍の速度）で特徴量抽出が可能であることが分かった（図 11）．MP3CEPについては 76.3%の高速化となった．

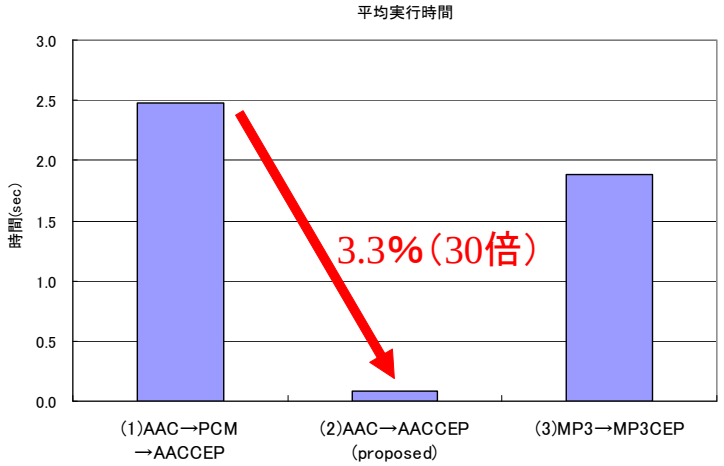


図 11　100 曲の楽曲データの特徴量抽出平均時間

4.3 精度検証

(1) **検証条件**

同様に検索精度に関しても検証を行った．検索条件は4.2と同様である．検索対象データ 100 曲について，クエリ楽曲（1 曲）からのコサイン距離を求めることを 10 回繰り返し，その結果の平均により，MFCCとの比較を行った．

(2) **検証結果**

図 12はMFCCによる検索結果で類似度の高い順にソートした結果である．順位の入れ替わりが少なく，滑らかな曲線に近い方が検索精度が高いと言える．数値的比較のため，MFCCとの相関値を求めた結果が表 2である．AACCEPでは従来のMP3CEPに比べて高精度な検索が行われている．

高域強調処理を入れた場合と入れない場合でも比較を行った結果，強調処理を行うことで精度が向上することが分かった．

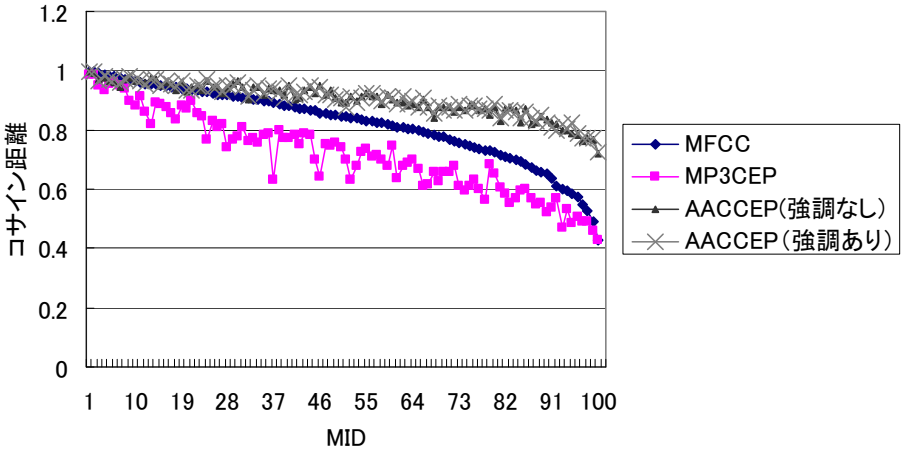


図 12　検索結果の比較

方式	相関値
MP3CEP	0.9389
AACCEP（強調なし）	0.9678
AACCEP（強調あり）	0.9730

表 2 MFCC との相関値

5. 考察

提案手法により，MFCC に近い検索結果を得ながら検索速度を約 30 倍に高速化することが可能となった．検索精度については，MFCC に近い検索結果を得ることを目標として，精度比較を行ったが，従来手法である MP3CEP よりも MFCC に近い値が得られることが分かった．また，MFCC で実施している高域強調についても取り入れることで，より MFCC に近い検索結果が得られることができた．

6. 応用例

本検索手法の応用例として，現在au oneラボ[7]で「にたうた検索」を公開中である．（図 13）．本システムは楽曲の類似検索の研究成果を直接一般ユーザに評価してもらうことを目的に開発された．ユーザは検索サーバにWebブラウザからアクセスし，選択した楽曲に近い楽曲を検索し，ダウンロード再生することができる．システム構築や更新の際，全楽曲の特徴量を予め用意する必要があるが，本提案手法を用いることでその高速化が可能となる．

7. おわりに

本稿では，AAC データを対象とした特徴量抽出の高速化手法について検討した．検索精度については，MFCC を正解データとして評価を行ったが，今後は主観的にも MFCC に近い，または MFCC よりも精度の高い楽曲検索結果が得られるよう改良を行う．

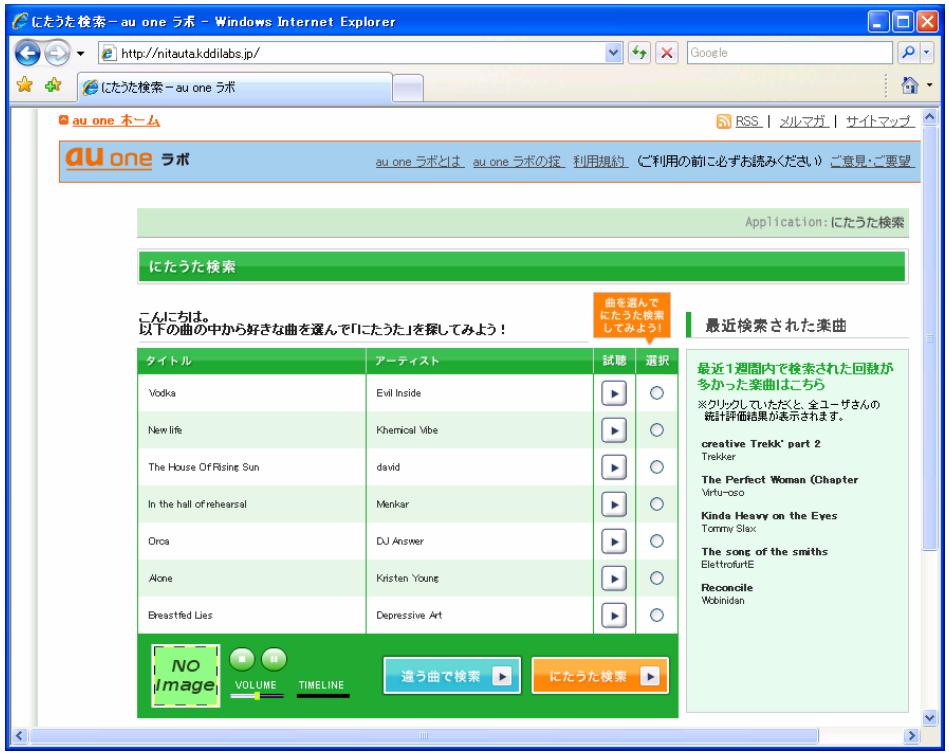


図 13　au one ラボで公開中の「にたうた検索」

参考文献

1 Keiichiro Hoashi et al., “FEATURE SPACE MODIFICATION FOR CONTENT-BASED MUSIC RETRIEVAL BASED ON USER PREFERENCES”,pp.517-520, ICASSP 2006.
2 ISO/IEC 11172-3:1993, “Information technology – Coding of moving pictures and associated audio for digital storage edia at up to about 1.5 Mbits/s – Part 3: Audio”, First edition, 1993.
3 ISO/IEC 14496-3:2005, "Information technology - Coding of audio-visual objects - Part 3: Audio", Third edition, 2005.
4 青木，神田，帆足，柳原，“楽曲類似検索における特徴量抽出の高速化”，情処第 71 回全大，2D-5, pp.2-43,44, 2009.
5 David Pye, “Content-Based Methods for the Management of Digital Music”, pp.2437-2440 vol.4, ICASSP 2000.
6 J. Foote, TreeQsoftware, http://treeq.sourceforge.net/
7 auoneラボ, http://lab.auone.jp/